

**REGRESI PROSES GAUSSIAN UNTUK PEMODELAN KALIBRASI  
SPEKTROSKOPI  
(STUDI KASUS : PENGUKURAN KONSENTRASI KURKUMIN, SEBUAH  
SENYAWA PENCIRI PADA TANAMAN OBAT TEMU LAWAK)**

**Moch. Abdul Mukid<sup>1</sup>, Aji Hamim Wigena<sup>2</sup>, Erfiani<sup>2</sup>**

1) Mahasiswa S2 Prodi Statistika IPB

2) Dosen Prodi Statistika Institut Pertanian Bogor

**Abstrak**

*Model-model kalibrasi multivariat telah dikembangkan dengan menggunakan metode regresi melalui pendekatan teknik regresi komponen utama dan kuadrat terkecil sebagian. Penelitian ini mengusulkan penerapan regresi proses gaussian sebagai metode alternatif. Sebuah proses gaussian diturunkan dari perspektif regresi nonparametrik bayesian dimana pendugaan nilai hyperparameternya dilakukan dengan metode kemungkinan maksimum. Untuk mengatasi banyaknya peubah bebas yang terlibat, pereduksian peubah dilakukan dengan metode analisis komponen utama. Regresi proses gaussian lebih fleksibel jika dibandingkan dengan metode-metode sebelumnya, dalam arti bahwa dengan pemilihan fungsi peragam yang tepat dia mampu menangkap struktur linear maupun nonlinear dari gugus-gugus data yang diteliti.*

**Kata kunci :** *Proses Gaussian , Regresi Proses Gaussian, Fungsi Peragam, Hiperparameter, Metode Kemungkinan Maksimum*

**PENDAHULUAN**

Sebagai alat analisis yang ampuh, teknik-teknik spektroskopi seperti spektroskopi massa dan infra merah, telah meningkat implementasinya di berbagai sektor seperti makanan, obat-obatan dan kimia. Spektrum yang dihasilkan berupa sebuah kurva kontinyu yang diukur dari ratusan atau bahkan ribuan panjang gelombang dan diasumsikan bahwa secara tidak langsung menangkap sifat-sifat fisis maupun kimiawi dari material yang sedang dianalisis (Chen *et al*, 2005). Analisis berikutnya dari spektrum dapat berupa kualitatif ataupun kuantitatif. Analisis kualitatif secara umum ditujukan untuk masalah-masalah klasifikasi, seperti misalnya klasifikasi dari kanker rahim dengan menggunakan spektroskopi massa; dan masalah-masalah pendeteksian, seperti misalnya deteksi proses transisi antar posisi steady states dalam penyaringan minyak. Sebaliknya, analisis kuantitatif fokus pada penentuan nilai dari sifat-sifat fisis maupun kimiawi dari spektrum yang diukur, seperti misalnya penentuan persentase berat dari substansi aktif dalam sebuah sample tablet. Hal ini dikenal sebagai kalibrasi.

Teknik-teknik kalibrasi tradisional didasarkan pada metodologi regresi linear.. Misal  $\mathbf{x}_i$  menandakan spectrum dari sampel ke- $i$  , yaitu sebuah vektor yang berisi nilai-nilai pada  $p$  panjang gelombang . Jika semua spectrum dari keseluruhan sampel digabung, diperoleh matrik input  $\mathbf{X} = (\mathbf{x}_1, \dots, \mathbf{x}_n)^T$  . Demikian juga, misal  $\mathbf{y}_i$  adalah sebuah vektor berdimensi  $q$ , dimana  $q$  adalah banyaknya output yang diperhatikan maka  $\mathbf{Y} = (\mathbf{y}_1, \dots, \mathbf{y}_n)^T$  . Tugas kalibrasi adalah untuk membangun sebuah model regresi multivariate dengan pendekatan matematis yaitu  $\mathbf{Y} = f(\mathbf{X})$ . Apabila  $f(\mathbf{X})$  adalah polynomial berderajat satu maka model yang bersesuaian adalah regresi linear multivariate, yaitu :

$$\mathbf{Y} = \mathbf{XB} + \mathbf{E} \quad (1)$$

Dimana  $\mathbf{B}$  adalah matrik koefisien regresi dan  $\mathbf{E}$  adalah matrik residual. Dengan menggunakan metode kuadrat terkecil, penduga matrik  $\mathbf{B}$  adalah

$$\mathbf{B} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y} \quad (2)$$

Salah satu masalah yang dihadapi dengan pendekatan ini adalah bahwa dalam beberapa situasi jumlah training sampel lebih kecil daripada jumlah predictor, sehingga akan menyebabkan matrik  $\mathbf{X}^T \mathbf{X}$  singular. Oleh karena itu tidak mungkin mendapatkan invers dari  $\mathbf{X}^T \mathbf{X}$  dan pada akhirnya koefisien regresi di persamaan (2) juga tidak diperoleh. Ini adalah situasi lumrah dalam pengembangan model kalibrasi dalam bidang spektroskopi dimana jumlah prediktor (panjang gelombang) lebih banyak dari jumlah sampel yang diamati.

Masalah ini diatasi dengan mengembangkan sebuah model regresi yang didasarkan atas pereduksian jumlah peubah penjelas menjadi peubah-peubah baru yang saling bebas antara satu dengan lainnya, diantaranya adalah regresi komponen utama dan regresi kuadrat terkecil sebagian.

Beberapa penulis telah mengembangkan model kalibrasi dengan mempertimbangkan adanya kondisi multikolinearitas diatas. Atok (2005) menggunakan Jaringan Syaraf Tiruan dengan metode pra pemrosesan Analisis Komponen Utama, sedangkan Djuraidah (2003) membandingkan performan model PLS non-linear dengan Jaringan Syaraf Tiruan pada model kalibrasi. Erfiani (2005) mengembangkan model kalibrasi dengan pendekatan bayes dimana reduksi peubahnya melalui pendekatan regresi terpenggal, sedangkan Sony (2005) menggunakan metode wavelet untuk tahap pra pemrosesannya. Selain pendekatan parametrik, beberapa penulis juga mengembangkan model kalibrasi dengan pendekatan non parametrik. Tonah (2005) menggunakan regresi sinyal P-Spline dengan metode pra pemrosesannya koreksi pencaran multiplikatif untuk kalibrasi data gingerol. Sedangkan Chen *et al* (2005) membandingkan performan Regresi Proses Gaussian (RPG) dengan metode regresi komponen utama, jaringan syaraf tiruan dan regresi kuadrat terkecil sebagian pada pemodelan kalibrasi untuk menentukan persentase bobot senyawa aktif pada tablet Escitalopram. Dalam pendugaan hiperparameternya mereka menggunakan simulasi Markov Chain Monte Carlo (MCMC) dengan algoritmanya Metropolis-Hasting. Akhirnya mereka menyimpulkan bahwa regresi proses gaussian menghasilkan RMSEP (*Root Mean Square Error Prediction*) yang lebih kecil dari ketiga metode tersebut. Termotivasi oleh kerja dari Chen dan kawan-kawan, maka penelitian ini bertujuan untuk memperoleh model kalibrasi terbaik pada pengukuran konsentrasi senyawa aktif kurkumin dengan menggunakan Regresi Proses Gaussian.

Pada bagian 2 akan dijelaskan secara ringkas tentang Proses Gaussian. Di bagian 3 dari tulisan ini, kami akan memaparkan lebih detail tentang Proses Gaussian untuk metodologi regresi. Karena dalam Regresi Proses Gaussian perlu membuat spesifikasi dari fungsi peragam, maka pada bagian 4 akan dijelaskan tentang definisi dan jenis-jenis fungsi peragam. Sedangkan pada bagian 5 nya kami akan membahas tentang bagaimana teknik menduga hiperparameter dari fungsi peragam tersebut. Bagian 6 dari tulisan ini berisi tentang ilustrasi penerapan Regresi Proses Gaussian untuk pemodelan spektroskopi. Akhirnya di bagian 7 kami akan memberikan kesimpulan atas penelitian ini.

## PEMBAHASAN

### Proses Gaussian

Proses stokastik adalah sebuah koleksi dari peubah-peubah acak  $\{Y_{\mathbf{x}} | \mathbf{x} \in X\}$  yang diindekskan oleh sebuah himpunan  $X$ . Dalam kasus ini,  $X$  biasanya berdimensi  $d$ , dimana  $d$  adalah jumlah input. Proses-proses stokastik di tentukan oleh pemberian sebaran peluang bersama untuk setiap himpunan bagian manapun dari  $Y_{\mathbf{x}_1}, \dots, Y_{\mathbf{x}_k}$  dengan sebuah cara yang konsisten. Proses Gaussian adalah sebuah proses stokastik yang mana himpunan berhingga manapun dari peubah  $Y$  mempunyai sebarang bersama Gaussian ganda. Sebuah proses Gaussian secara lengkap ditentukan oleh fungsi rata-rata  $\mu(\mathbf{x}) = E[Y_{\mathbf{x}}]$  dan fungsi peragam  $k(\mathbf{x}_i, \mathbf{x}_j) = E[(Y_{\mathbf{x}_i} - \mu(\mathbf{x}_i))(Y_{\mathbf{x}_j} - \mu(\mathbf{x}_j))]$  (Williams, 2002). Untuk ruang input multidimensi, proses Gaussian disebut dengan medan peubah acak Gaussian.

Rasmussen dan Williams (2006) mendefinisikan proses Gaussian sebagai sebuah kumpulan dari peubah-peubah acak dimana sebanyak berhingga dari peubah-peubah tersebut memiliki sebaran bersama Gaussian. Proses Gaussian dispesifikasikan secara lengkap oleh fungsi rata-rata dan fungsi peragamnya. Jika  $m(\mathbf{x})$  sebagai fungsi rata-rata dan  $k(\mathbf{x}_i, \mathbf{x}_j)$  sebagai fungsi peragam dari sebuah Proses Gaussian  $f(\mathbf{x})$  maka didefinisikan :

$$m(\mathbf{x}) = E [ f(\mathbf{x}) ] \quad (3)$$

$$k(\mathbf{x}_i, \mathbf{x}_j) = E [ (f(\mathbf{x}_i) - m(\mathbf{x})) (f(\mathbf{x}_j) - m(\mathbf{x})) ] \quad (4)$$

dan Proses Gaussian ditulis dengan

$$f \approx GP ( m(\mathbf{x}), k(\mathbf{x}_i, \mathbf{x}_j) ) \quad (5)$$

artinya : “ fungsi  $f$  tersebar sebagai sebuah Proses Gaussian dengan fungsi rata-rata  $m(\mathbf{x})$  dan fungsi peragam  $k(\mathbf{x}_i, \mathbf{x}_j)$ ”.

Sebagai teladan dari proses Gaussian adalah regresi linier bayes  $y = \mathbf{x}^T \boldsymbol{\beta}$ , dimana  $\mathbf{x}$  adalah vector input dan  $\boldsymbol{\beta}$  adalah koefisien regresinya. Misal  $\boldsymbol{\beta}$  dianggap sebagai peubah acak yang mempunyai sebaran Gaussian dengan rata-rata  $\mathbf{0}$  dan matrik peragam  $\boldsymbol{\Sigma}$ , maka  $\mu(\mathbf{y}) = 0$  dan

$$\text{kov}(\mathbf{y}_i, \mathbf{y}_j) = E[(Y_{x_i} - \mu(\mathbf{x}_i))(Y_{x_j} - \mu(\mathbf{x}_j))] = E[Y_{x_i} Y_{x_j}] = \mathbf{x}_i^T \boldsymbol{\Sigma} \mathbf{x}_j \quad .$$

### Regresi Proses Gaussian

Pada bagian ini akan dijelaskan sebuah model dimana diasumsikan bahwa masing-masing amatan  $y_i$  bergantung pada sebuah peubah laten  $f_i$  sebagai berikut :

$$y_i = f(\mathbf{x}_i) + \varepsilon_i \quad (6)$$

dimana  $\varepsilon_i$  adalah peubah acak galat yang secara bebas dan identik menyebar Gaussian dengan rata-rata nol dan ragam  $\sigma^2 \mathbf{I}$ , sedangkan  $\mathbf{x}_i$  adalah vektor input ke-i. Apabila peubah laten yang diperhatikan sebanyak  $n$  untuk kemudian dikumpulkan dalam sebuah vektor  $\mathbf{f} = [f_1 \dots f_n]^T$  maka menurut Proses Gaussian untuk metodologi regresi, sebaran prior atas vektor  $\mathbf{f}$  adalah Gaussian Ganda dengan vektor rata-rata nol dan matrik peragam  $\mathbf{K}$ , yaitu

$$\mathbf{f} | \mathbf{X}, \boldsymbol{\theta} \approx N(\mathbf{0}, \mathbf{K}) \quad (7)$$

dimana  $\mathbf{K}$  adalah matrik  $n \times n$  yang bergantung pada  $\mathbf{X}$  dan beberapa parameter  $\boldsymbol{\theta}$ . Setiap elemen ke (i,j) dari matrik  $\mathbf{K}$  adalah sama dengan  $k(\mathbf{x}_i, \mathbf{x}_j)$  dimana  $k(\dots)$  adalah sebuah fungsi yang definit positif yang terparameterisasi oleh  $\boldsymbol{\theta}$ . Dalam konteks ini  $k(\dots)$  disebut sebagai sebuah fungsi peragam.

Misal diberikan beberapa amatan dan sebuah fungsi peragam, kemudian akan ditentukan sebuah prediksi dengan menggunakan Model Proses Gaussian. Untuk melakukan hal itu, jika  $\mathbf{x}^*$  sebuah titik uji dan  $f^*$  adalah peubah laten yang bersesuaian, maka dibawah kerangka kerja Proses Gaussian, sebaran bersama dari  $\mathbf{f}$  dan  $f^*$  adalah Gaussian Ganda dengan rata-rata nol dan diperoleh melalui menambah persamaan (5) dengan peubah laten baru  $f^*$  :

$$\begin{bmatrix} \mathbf{f} \\ f^* \end{bmatrix} | \mathbf{X}, \boldsymbol{\theta} \approx N \left( \mathbf{0}, \begin{bmatrix} \mathbf{K} & \mathbf{k} \\ \mathbf{k}^T & \kappa \end{bmatrix} \right) \quad (8)$$

dimana  $\mathbf{k} = [k(\mathbf{x}^*, \mathbf{x}_1), \dots, k(\mathbf{x}^*, \mathbf{x}_n)]^T$  adalah vektor  $n \times 1$  yang dibentuk dari peragam antara  $\mathbf{x}^*$  dan input-input training. Sedangkan  $\kappa = k(\mathbf{x}^*, \mathbf{x}^*)$  adalah sebuah skalar. Apabila peubah galat mengikuti sebaran seperti pada persamaan (6) maka sebaran bersama dari peubah teramati  $\mathbf{y}$  dan  $y^*$  adalah

$$\begin{bmatrix} \mathbf{y} \\ y^* \end{bmatrix} | \mathbf{X}, \boldsymbol{\theta} \approx N \left( \mathbf{0}, \begin{bmatrix} \mathbf{K} + \sigma^2 \mathbf{I} & \mathbf{k} \\ \mathbf{k}^T & \kappa + \sigma^2 \end{bmatrix} \right) \quad (9)$$

Sehingga  $y^*$  mengikuti sebaran normal juga, yaitu :

$$y^* | \mathbf{y}, \mathbf{X}, \boldsymbol{\theta} \approx N(m(\mathbf{x}^*), v(\mathbf{x}^*)) \quad (10)$$

dimana rata-rata dan ragam prediktifnya adalah

$$m(\mathbf{x}^*) = \mathbf{k}^T (\mathbf{K} + \sigma^2 \mathbf{I})^{-1} \mathbf{y} \quad (11)$$

$$v(\mathbf{x}^*) = \kappa + \sigma^2 - \mathbf{k}^T (\mathbf{K} + \sigma^2 \mathbf{I})^{-1} \mathbf{k} \quad (12)$$

Apabila fungsi peragam yang didefinisikan oleh hyperparameters  $\boldsymbol{\theta}$  diketahui maka dapat dihitung sebuah sebaran prediktif Gaussian untuk sembarang titik uji  $\mathbf{x}^*$ . Nilai dugaan bagi  $y^*$  adalah  $m(\mathbf{x}^*)$  dan ragam bagi dugaan  $y^*$  adalah  $v(\mathbf{x}^*)$ . Secara umum untuk  $m$  buah titik uji  $\mathbf{X}^* = [\mathbf{x}_1^*, \dots, \mathbf{x}_m^*]$  maka sebaran prediktifnya Gaussian Ganda dengan parameter-parameter,

$$m(\mathbf{X}^*) = \mathbf{K}^{*T} (\mathbf{K} + \sigma^2 \mathbf{I})^{-1} \mathbf{y} \quad (13)$$

$$v(\mathbf{X}^*) = \mathbf{K}^{**} + \sigma^2 - \mathbf{K}^{*T} (\mathbf{K} + \sigma^2 \mathbf{I})^{-1} \mathbf{K}^* \quad (14)$$

dimana  $\mathbf{K}^*$  adalah matrik  $n \times m$  dari peragam antara input-input training dan titik-titik uji. Matrik  $\mathbf{K}^{**}$  dengan ukuran  $m \times m$  tersusun dari peragam antara titik-titik uji.

Selanjutnya akan dikenalkan fungsi kemungkinan marginal  $p(y|X)$ . Kemungkinan marginal adalah integral dari fungsi kemungkinan dikalikan dengan sebaran prior atas  $\mathbf{f}$ , yaitu

$$p(y|X) = \int p(y|\mathbf{f}, X) p(\mathbf{f}|X) d\mathbf{f} \quad (15)$$

Dibawah model proses gaussian sebaran priornya adalah gaussian, yaitu  $\mathbf{f}|X \approx N(\mathbf{0}, \mathbf{K})$  atau

$$\log p(\mathbf{f}|X) = -\frac{1}{2} \mathbf{f}^T \mathbf{K}^{-1} \mathbf{f} - \frac{1}{2} \log |\mathbf{K}| - \frac{n}{2} \log 2\pi \quad (16)$$

Sehingga fungsi kemungkinan marginalnya menurut Rasmussen (2006) adalah

$$\log p(\mathbf{y}|\mathbf{X}) = -\frac{1}{2} \mathbf{y}^T (\mathbf{K} + \sigma^2 \mathbf{I})^{-1} \mathbf{y} - \frac{1}{2} \log |\mathbf{K} + \sigma^2 \mathbf{I}| - \frac{n}{2} \log 2\pi \quad (17)$$

Persamaan (17) akan digunakan untuk menduga nilai-nilai hiperparamater dari fungsi peragam pada regresi proses gaussian.

### Fungsi Peragam

Fungsi peragam adalah sebuah fungsi dari input-input model yang hasilnya adalah nilai peragam untuk output-output yang bersesuaian (Rasmussen, 1996). Terdapat beberapa kemungkinan pilihan terhadap fungsi peragam. Secara formal, perlu menentukan sebuah fungsi dimana akan membangkitkan sebuah matrik ragam peragam yang definit non negatif untuk sembarang himpunan titik-titik input. Fungsi peragam ini memuat parameter-parameter yang nilainya perlu ditentukan. Parameter-parameter ini disebut dengan hiperparameter.

Seperti halnya dalam kasus pemodelan bayesian, nilai hiperparameter tidak ditentukan diawal, tetapi akan diadaptasikan ketika model cocok dengan struktur data. Secara garis besar fungsi peragam dapat dibedakan menjadi dua, yaitu fungsi peragam yang stasioner dan fungsi

peragam yang tidak stasioner. Fungsi ragam yang stasioner adalah sebuah fungsi dari  $\mathbf{x}_i - \mathbf{x}_j$  dimana  $\mathbf{x}_i$  dan  $\mathbf{x}_j$  adalah dua buah input data yang berbeda. Fungsi ragam yang stasioner invarian terhadap translasi namun seringkali gagal untuk menyesuaikan terhadap kemulusan dalam fungsi yang diteliti. Sebaliknya fungsi ragam yang tidak stasioner adalah bukan merupakan fungsi dari  $\mathbf{x}_i - \mathbf{x}_j$ , namun fungsi ragam jenis ini mampu menyesuaikan terhadap kemulusan fungsi (Paciorek & Schervish, 2005). Yang termasuk fungsi ragam yang stasioner adalah, kuadrat eksponensial,  $\gamma$ -eksponensial, dan kuadrat rasional. Sedangkan yang termasuk fungsi ragam tidak stasioner adalah fungsi-fungsi ragam linear dan neural network.

### Pendugaan Hiperparameter

Jika fungsi ragam sudah diketahui, maka ia dapat digunakan untuk membuat sebuah prediksi secara langsung dari titik uji baru. Namun dalam situasi-situasi praktis, fungsi ragam mana yang harus dipilih tidak diketahui. Salah satu opsinya adalah dengan memilih sebuah keluarga parametrik dari fungsi-fungsi ragam (dengan vektor parameter  $\boldsymbol{\theta}$  atau disebut juga dengan hiperparameter) dan kemudian menentukan nilai-nilai hiperparameter yang memberikan prediksi terbaik.

Ada beberapa metode untuk menentukan nilai  $\boldsymbol{\theta}$ . Williams (2002) menyatakan bahwa untuk menduga nilai  $\boldsymbol{\theta}$  dapat digunakan Metode Kemungkinan Maksimum, Metode Aposterior Maksimum dan metode simulasi Hybrid Monte Carlo. Metode lain yang bisa digunakan adalah metode Cross Validation dan metode Generalized Cross Validation (Wahba, 1990). Sedangkan Chen *et al* (2005) menggunakan algoritma Metropolis-Hasting untuk menduga nilai-nilai hiperparameter pada penelitiannya.

Dalam penelitian ini, pendugaan nilai hiperparameternya menggunakan metode kemungkinan maksimum. Penyesuaian  $\boldsymbol{\theta}$  difasilitasi oleh kenyataan bahwa fungsi kemungkinan marginal  $l = \log p(\mathbf{y}|\mathbf{X})$  dapat dihitung secara analitik, seperti pada persamaan (15). Pemaksimum fungsi kemungkinan marginal  $l$  dilakukan dengan metode Conjugate Gradien, yaitu salah satu metode optimasi untuk fungsi dari banyak peubah (Fletcher & Reeves, 1964). Rasmussen (1996) telah mengembangkan kode-kode dalam bahasa pemrograman matlab untuk operasionalisasi metode conjugate gradient ini. Oleh sebab itu dengan menyediakan argument-argumen tertentu, kode-kode tersebut dimanfaatkan sehingga mempermudah penerapan regresi proses gaussian untuk pemodelan kalibrasi.

### Data dan Analisis

Data yang digunakan adalah data hasil pengukuran HPLC dan FTIR pada ekstrak temu lawak. Hasil pengukuran HPLC berupa konsentrasi kurkumin dan hasil pengukuran FTIR berupa persen transmittansi dari kurkumin. Temu lawak yang dijadikan sampel berasal dari beberapa daerah sentra tanam obat, yaitu Bogor, Sukabumi, Semarang, Karanganyar, Yogyakarta dan Malang. Pengukuran dilakukan di laboratorium BIOFARMAKA, Institut Pertanian Bogor dengan jumlah sampel yang diamati sebanyak 20 buah.

Karena persen transmittansi diukur pada 1886 bilangan gelombang dan dalam konteks pemodelan kalibrasi ini berperan sebagai peubah penjelas, maka untuk menghindari munculnya masalah multikolinearitas, peubah ini perlu direduksi menjadi peubah-peubah baru yang tidak berhubungan antara satu dengan yang lainnya. Metode yang kami gunakan adalah dengan menggunakan Analisis Komponen Utama dimana kriteria untuk menentukan banyaknya komponen utama adalah persentase keragaman yang mampu dijelaskan. Dari Tabel 1 tampak bahwa dengan menggunakan 3 komponen keragaman yang dapat dijelaskan sebesar 99,266% dari keragaman pada data asal.

Tabel 1 : Ragam kumulatif untuk banyak komponen utama yang akan digunakan (Data : konsentrasi kurkumin, senyawa aktif pada temu lawak)

Jml Komponen	Ragam yg dijelaskan	Proporsi ragam (%)	Ragam Kumulatif (%)
1	29,578	90,089	90,089
2	1,722	5,244	95,333
3	1,291	3,933	99,266
4	0,156	0,474	99,740
5	0,039	0,117	99,858

Dengan menggunakan skor dari 3 komponen utama ini, dan menganggap hasil pengukuran dari HPLC sebagai peubah respon maka pemodelan kalibrasi dengan menggunakan regresi proses gaussian dapat dilakukan. Untuk memenuhi asumsi bahwa peubah respon harus menyebar gaussian ganda dengan nilai tengah nol, maka hasil pengukuran dari HPLC terlebih dahulu dikurangi dengan rata-ratanya.

Pengolahan data dilakukan dengan memanfaatkan ToolBox gpml dalam bahasa pemrograman matlab yang dibuat oleh Rasmussen (1996). ToolBox tersebut dapat diunduh secara bebas di situs <http://www.gaussianprocess.org>. Berikut ini adalah nilai KTG dari regresi proses gaussian untuk berbagai fungsi peragam.

Tabel 2 Nilai KTG regresi proses gaussian dengan berbagai fungsi peragam

(Data : konsentrasi kurkumin, senyawa aktif pada temu lawak)

Kategori	Fungsi Peragam	KTG
Stasioner	Kuadrat Eksponensial (isotropik)	0,0275
	Kuadrat Eksponensial (ARD)	<b>0,0168</b>
Non Stasioner	Linear (ARD)	0,1728
	Linear (One)	0,2323
Gabungan	Linear ARD + KE-ARD	0,0172
	Linear -1 + KE-ARD	<b>0,0168</b>
Regresi Komponen Utama		0,2013

Dari hasil diatas (Tabel 2), dapat disimpulkan bahwa dengan menggunakan indikator nilai KTG, secara umum regresi proses gaussian memberikan kinerja yang lebih baik jika dibandingkan

dengan regresi komponen utama. Dapat disimpulkan pula bahwa fungsi peragam yang relevan untuk model kalibrasi ini adalah kuadrat eksponensial-ARD, karena menghasilkan nilai KTG yang paling kecil, yaitu 0,0168. Meskipun dengan menggunakan fungsi peragam gabungan antara Linear-1 dan Kuadrat eksponensial-ARD juga menghasilkan nilai KTG yang sama, namun pencapaian konvergensi nilai hiperparameter jauh lebih lama bila dibandingkan dengan menggunakan fungsi peragam kuadrat eksponensial saja. Selain itu dari Tabel 2 diatas, tampak bahwa penambahan suku linear-1 pada kuadrat eksponensial – ARD tidak mampu memperkecil nilai KTG.

## KESIMPULAN

Dengan dukungan dari kemajuan komputasi kestatistikan, regresi proses gaussian makin mudah untuk diimplementasikan dalam pemodelan kalibrasi spektroskopi. Kesulitan-kesulitan yang berkaitan dengan optimasi dari fungsi banyak peubah dan menentukan invers dari matrik yang berukuran besar, telah dapat diatasi dengan pengembangan kode-kode dalam berbagai bahasa pemrograman, khususnya MATLAB. Pada penelitian ini dapat disimpulkan bahwa fungsi peragam yang relevan untuk pemodelan kalibrasi pada pengukuran konsentrasi kurkumin adalah Kuadrat Eksponensial – ARD.

## DAFTAR PUSTAKA

- Atok RM. 2005. Jaringan Syaraf Tiruan dalam Pemodelan Kalibrasi dengan Pra-pemrosesan Analisis Komponen Utama dan Transformasi Fourier Diskrit [Tesis]. Bogor: Program Pascasarjana, Institut Pertanian Bogor.
- Chen T, Morris J, Martin E. 2005. *Gaussian Process Regression for Multivariate Spectroscopic Calibration*. Newcastle: School of Chemical Engineering and Advanced Material.
- Djuraidah RM. 2003. Penerapan Model Nonlinear PLS dengan Jaringan Syaraf Tiruan dalam Kalibrasi. *Jurnal Matematika Aplikasi dan Pembelajarannya (JMAP)* 2:339-345. Bogor: Program Pascasarjana, Institut Pertanian Bogor.
- Erfiani. 2005. Pengembangan Model Kalibrasi dengan Pendekatan Bayes (Kasus Tanaman Obat [Disertasi]. Bogor: Program Pascasarjana, Institut Pertanian Bogor.
- Danutirto H. 2001. Pengembangan fitofarmaka di Indonesia. Lokakarya dan pameran Pengembangan Agribisnis Berbasis Biofarmaka. Kerjasama Departemen Pertanian dengan Institut Pertanian bogor. Jakarta Tgl 13-16 November 2001.
- Fletcher, R. and Reeves, C. M. (1964) Function minimization by conjugate gradients. *Computer Journal* 7, 148–154.
- Naes T, Issackson T, Fearn T, Davies T. 2002. *Multivariate Calibration and Classification*. United Kingdom: NIR Publication Chichester.
- Nur MA, Adjuwana H. 1992. *Teknik Spektroskopi dalam Analisis Biologi*. Bogor: Pusat Antar Universitas, Institut Pertanian Bogor.
- Paciorek CJ, Schervish M. 2005. *Nonstationary Covariance Functions for Gaussian Process Regression*. Pittsburgh : Departmet of Statistics, Carnegie Mellon University.
- Rasmussen CE. 1996. [Evaluation of Gaussian Processes and other Methods for Non-linear Regression](#) [Disertasi]. Toronto: Department of Computer Science, University of Toronto.
- Rasmussen CE, Williams CKI. 2006. *Gaussisn Process for Machine Learning*. Massachusetts : MIT Press.

- Sunaryo S. 2005. Model Kalibrasi dengan Transformasi Wavelet sebagai Metode Pra-pemrosesan [Disertasi]. Bogor: Program Pascasarjana, Institut Pertanian Bogor.
- Tonah. 2006. Pemodelan Kalibrasi Peubah Ganda dengan Pendekatan Regresi Sinyal P-Spline [Tesis]. Bogor: Program Pascasarjana, Institut Pertanian Bogor.
- Williams CKI. 2002. [Gaussian processes](#). Di dalam: M. A. Arbib, editor. *Handbook of Brain Theory and Neural Networks*. Ed ke-2. Cambridge: MIT Press. hlm 466-470.